A complementary analysis for SAGE II data profiles

D. Rind,[1] J. Lerner,[2] and J. Zawodny[3]

Received 25 January 2005; revised 4 March 2005; accepted 16 March 2005; published 14 April 2005.

[1] We present a screened and gridded SAGE II data set, as a complement to the profile data available for the 19 years of records from November 1984 through December 2003. The data screening method is described, and differences between screened and unscreened data are presented. Extensive changes occur in the water vapor data set, which has excessive values in many regions without the screening. Screened NO_2 values are larger in the middle stratosphere, for the zonal average by up to 10–15%, and reduced in the lower stratosphere zonally by up to 15–30%, while zonal average ozone values are as much as 20–30% larger in the lower stratosphere and upper troposphere. Many of these changes are due to interference effects from aerosols and clouds, but additional bad data points, isolated by visual inspection, occur for no obvious reason. The screening procedures also remove spurious trends. This data set is available via CD-ROM. **Citation:** Rind, D., J. Lerner, and J. Zawodny (2005), A complementary analysis for SAGE II data profiles, *Geophys. Res. Lett.*, *32*, L07812, doi:10.1029/2005GL022550.

1. Introduction

[2] SAGE II was launched in November 1984 [*McCormick*, 1987], and has been collecting data on various atmospheric species since that time, with a few exceptions. The level 2 data is available from the NASA Langley Radiation and Aerosols Branch in the form of profiles of various species. The data is purposely not screened, in an effort to provide researchers with the maximum amount possible independent of interpretation. Some flags are provided to indicate potentially unreliable data, and nominal error bars are included.

[3] As a complement to that data set, and to aid in data utilization, we have screened the Version 6.2 data to remove data that is deemed inappropriate (as defined below). The screening procedure is done first on the individual raw profiles. Then each profile is assigned to a grid box of 7.8° latitude × 10° longitude, and all observations within the grid box are averaged to produce the monthly mean. Results are initially tabulated at 0.5 km vertical resolution, and then interpolated to 16 standard constant pressure levels. This manuscript discusses the approach utilized, presents comparisons between the screened and unscreened results, and advertises the availability of the resultant product.

[4] Screening data is always a tricky procedure; the experience in which the ozone hole was missed in satellite data because of screening of unexpected low values will forever remain a cautionary tale. However, clearly some data is inappropriate, either for reasons of physics (i.e., unrealistic supersaturations), unforeseen instrument problems, or obscuration by aerosols. The occultation approach data sets also risk having retrieval errors made at higher altitudes (where these problems exist) contaminate data at lower elevations (where they may not). Therefore we have used a variety of approaches to examine and filter the data product, including data analysis via means and standard deviations, augmented by visual inspection. The data screening procedures are discussed below.

2. Data Screening and Results

[5] The SAGE II data products consist of aerosol extinction at 4 wavelengths (1020NM, 525NM, 452NM, 386NM), O_3. H_2O, and NO_2. The National Center for Environmental Prediction (NCEP) provides temperature, geopotential height and tropopause pressure for each SAGE II retrieval. Together the two data sets allow for some subsidiary data products (relative humidity, integrated stratospheric aerosol extinction above the tropopause), and data analysis adds an additional quantity (cloud cover). Thirty profiles are collected each day, therefore there are some 900 profiles per month, and in excess of 200,000 profile opportunities over the 19-year time period. Close to 80% of these opportunities (~160,000) ultimately contained usable data. Data screening was applied to the species data (O_3. H_2O, and NO_2) in the individual profiles, before either the gridding or interpolation to constant pressure level.

[6] The screening was conducted according to the following criteria. (a) Water vapor: This data product required the greatest amount of screening, including physically implausible values (relative humidity $\gg$ 100% with respect to water, or <0%). The chief cause is contamination by aerosols. The 940NM water vapor channel is weak, and easily obscured by aerosol scattering. Since there is no aerosol channel in close proximity, aerosol effects at this wavelength are estimated from those at other wavelengths, introducing uncertainty into the procedure. When aerosol loading is small, the effect is minimal, but often that is not the case (e.g., after major volcanoes, or in the presence of clouds). The screening procedure removed data in the first 6 km when there was large absorption (integrated value above that level > about 5%) in the 1020NM channel (bit 5 true), when there were clouds between 6 and 25.5 km (bits 11, 12 true) and above the tropopause when the 1020NM absorption was greater than 4×10^{-4} km^{-1}. The last prescription is close to the screening criteria used by *Thomason et al.* [2004]. The presence of clouds follows the algorithm developed by *Kent et al.* [1995].

[1]NASA Goddard Institute for Space Studies, New York, New York, USA.
[2]Center for Climate Systems Research, Columbia University, New York, New York, USA.
[3]NASA Langley Research Center, Hampton, Virginia, USA.

0094-8276/05/2005GL022550$05.00

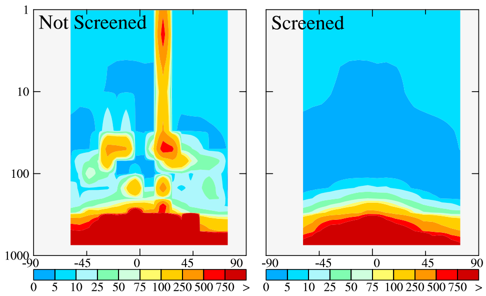

Figure 1. Climatological April water vapor as a function of latitude and pressure, (left) unscreened, (right) screened (ppmv).

[7] In addition, there were time periods with bad data (6/23/93–4/10/94 at pressures $\leq$30 mb; 11/91, 9/92, 10/92 at pressures $\geq$150 mb). As noted by *Wang et al.* [2002], between mid-1993 and mid-1994, SAGE II had a battery problem; so to conserve power, sunset measurements were started later than normal, while sunrise events were ended earlier. These events were thus much shorter than usual, reducing the extraterrestrial solar irradiance measurements required for normalization. They affected all three of the gases discussed here.

[8] In other cases, some bad events were noted at certain pressure levels, for no obvious physical reason. They were flagged through the use of movies (latitude $\times$ altitude, latitude $\times$ longitude) which revealed individual points that were deemed inappropriate, e.g., isolated values > ten times the local standard deviation that would appear and then disappear by the next retrieval. Again, without any specific cause, removing data is risky, but retaining data which is many standard deviations away from any other observation skews the resulting climatologies. This procedure suggested an additional 594 bad points.

[9] What is the effect of screening this data? Shown in Figure 1 are the screened and unscreened data sets for the April climatology. Removal of the various forms of bad data produces a *usable* water vapor climatology, removing in particular extremely large values. As shown in Figure 2, both the mean and (interannual) standard deviation have much less noise.

[10] (b) NO_2: Data were removed when the 1020NM absorption >7 $\times$ 10^{-4} km^{-1}, and again for specific events at higher altitudes. In addition, as suggested by the movies, large values, many times greater in magnitude than other data for a particular month and level, were removed; invariably they were there only for one retrieval. Excessively low values also occurred occasionally in conjunction with large aerosol absorption.

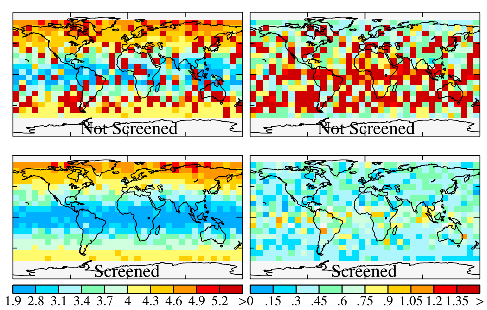

Figure 2. (top) Unscreened and (bottom) screened water vapor (ppmv) at 70 mb, averaged over all April data between 1985 and 2003. Shown are (left) the mean values and (right) the standard deviations of all the data points.

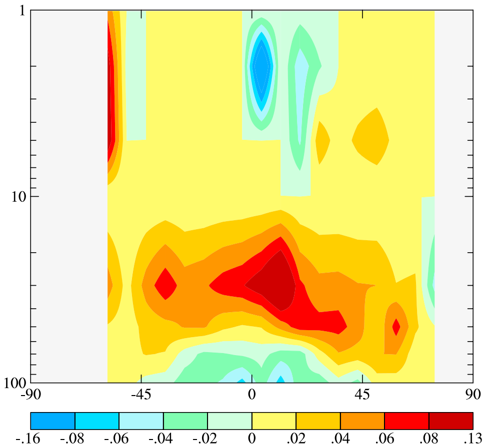

Figure 3. Difference between screened and unscreened April climatology for NO_2 (ppbv).

[11] An example of the difference between the screened and unscreened data is shown for the climatological April data in Figure 3. The screening results in higher NO_2 values in the middle stratosphere, increases on the order of 10–15%, and lower values in the lower stratosphere, of order 15–30%. Changes at individual latitude $\times$ longitude locations are of course at times much larger.

[12] (c) Ozone: Ozone had less necessity for screening, although large values were again removed following visual inspection (i.e. 10.5–24.5 km, values > 10 ppmv; above 25 km, values > 100 ppmv; at pressure <3 mb, values > 50 ppmv). Data were also removed at the lowest levels when the 1020NM absorption was large, in the mid-to-upper troposphere when clouds were detected, and for specific events at the highest levels.

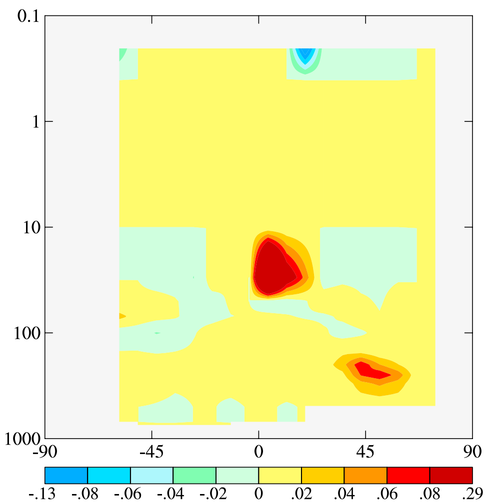

Figure 4. As in Figure 3 except for O_3 (ppmv).

Table 1. Percentage of Data Retained (Out of 159,113 Profiles)

Pressure, MB	H_2O	NO_2	O_3
700	14.4	NA	3.9
500	38.0	NA	28.6
300	43.9	NA	46.5
250	41.1	NA	56.4
200	47.5	NA	69.3
150	57.3	NA	80.5
100	59.4	64.1	86.0
70	64.6	76.8	91.0
50	76.4	83.4	92.2
30	84.8	91.8	95.6
10	93.5	98.5	98.7
5	94.1	93.8	98.8
2	93.8	81.5	98.7
1	87.1	47.7	98.6
0.4	NA	NA	98.6
0.2	NA	NA	91.0

[13] *Wang et al.* [2002] suggested screening ozone data based on aerosol extinctions of 6×10^{-3}/km, as well as extinction ratios of 525/1020 nm <1.4; the goal was to remove the effect of both volcanic aerosol and cloud interference. Our tests did not clearly indicate data that was always significantly different under these circumstances; thus we chose to remove anomalous data for ozone primarily when the resultant value was far removed from the mean and standard deviation of the climatological data set. However, almost all of the Pinatubo time period was removed because the aerosol effect was so strong that the cloud characterization mistook the aerosol for cloud.

[14] The difference between screened and unscreened values is shown in Figure 4. Again the screening (primarily for aerosols) has resulted in higher values, on the order of 20–30% in the lower stratosphere and upper troposphere; Wang et al. discuss the low ozone values following the Pinatubo eruption, relating to the aerosol size distribution. Again changes at individual locations can be much larger.

[15] How much of the data was ultimately screened by the various tests? Close to 160,000 usable profiles are included in the 19 years of SAGE II data. Given in Table 1 are the percentage of successful retrievals (i.e., those that passed the various screening criteria). Throughout most of the stratosphere, ozone data passed the screening criteria greater than 90% of the time, and for H_2O and O_3, screened retrievals even at the 300 mb level were achieved some 40% of the time (data retrievals for NO_2 stop in the lower stratosphere region just above the tropopause). Note that in some cases the data were unavailable even before the screening procedure was applied (i.e., the retrieval did not get to the lower level because of obscuration by cloud cover), an effect which is included in the table.

[16] SAGE data is often used for trend assessment, due to its self-calibration characteristics and endurance. We have calculated how the screening alters the trend analyses for these species. In the case of water vapor, the presence of aerosols, particularly Pinatubo-related, induced an artificially positive trend early in the record, with a subsequent negative trend; in the screened data, there is no early peak. [It should be noted, that because of instrument wavelength shifts during the record, the SAGE II data includes an advisory against using the data set for water vapor trends.] For NO_2, the unscreened data shows trends that differ with altitude and also between the early and latter part of the record (as might be envisioned by noting the corrections in Figure 3). In the unscreened data, many of those trends are removed or greatly ameliorated. Ozone variations are dominated above about 20 mb by an apparent solar cycle effect (i.e., 11 year variations) which are somewhat more distinct in the screened data set, as Pinatubo induced low values near the otherwise solar-maximum period (opposite to the correction in Figure 4). Spurious trends in the lower stratosphere are also no longer evident.

[17] As examples, we show in Table 2, the mean (linear) trend for the 19 years of the data at several different pressure levels for both the screened and unscreened data. As the latitudinal coverage varies monthly, (area-weighted) trends were calculated separately for each month, and then averaged to produce the yearly value. At levels below 10 mb the screening produces differences for each of the species, while in the upper stratosphere, water vapor is most affected.

3. SAGE CD

[18] This complementary data set is being made available on a CD-ROM. It is the third in a series of CDs that provides gridded, averaged data from the individual SAGE profiles. The first CD, distributed through the Langley DAAC, provided GIFs and gridded binary data for all the SAGE II standard retrievals (4 aerosol channels, ozone, water vapor, NO_2) as well as NMC-generated correlative data (temperature, geopotential height, tropopause pressure) for each month of the SAGE II retrieval from January of 1985 through 1993 (1991 for water vapor) [*Rind and Liao*, 1997]. It utilized SAGE V5.93 (5.91 for water vapor). A second CD provided the climatological mean values of the same data products plus relative humidity from 1985 to the Pinatubo eruption (June 1991), as well as the standard deviations and number of observations. Again it included both visual products and gridded data, and was distributed via both the SAGE II/III team (Pat McCormick, PI) and from GISS. This newest version is similar to Version 2, but utilizes SAGE V6.20, extends the data through December 2003, and adds cloud cover and additional presentations (global changes as a function of time, sunrise/sunset for NO_2, monthly data) as well as more rigorous screening.

Table 2. Linear Trends for the 19 Years of SAGE II Data, Both Screened and Unscreened

	H_2O, ppmv/yr		NO_2, ppbv/yr		O_3, ppbv/yr	
	Unscreened	Screened	Unscreened	Screened	Unscreened	Screened
100 mb	−25.778	0.003	−0.005	−0.002	0.631	0.698
30 mb	−6.460	−0.023	0.018	0.015	−2.649	−4.062
5 mb	0.005	0.005	0.047	0.047	−11.868	−11.883
1 mb	−0.026	0.056	0.023	0.025	4.594	4.679

[19] The CD contains the climatology and standard deviations for SAGE data products for the full 19 year time period plus the data (lat × long and lat × pressure) for each month of the time period. The climatology is presented as both binary data and images (both GIF and PostScript). The presentation is in three forms: Global average values as f(time) × pressure; lat × pressure; and lat × long (700, 500, 300, 250, 200, 150, 100, 70, 50, 30, 10, 5, 2, 1, 0.4, 0.2 mb). Monthly averages are provided, along with the standard deviations (for at least three events), obtained from the total number of data points for the 19 year time period (for that month) and the distribution and number of the gridded data points. The CD is available by sending an email to drind@giss.nasa.gov, with the title SAGE CD. The gridded data should be especially applicable for use by modelers. For complementary (screened) images – seasonal climatologies, and individual months for each year – see our website http://www.giss.nasa.gov/data/sageii/.

[20] **Acknowledgment.** Funding for SAGE II and related data products comes from the NASA ACMAP program/Atmospheric Concentration focus area.

References

Kent, G. S., P.-H. Wang, M. P. McCormick, and K. M. Skeens (1995), Multiyear SAGE II measurements of upper tropospheric aerosol characteristics, *J. Geophys. Res.*, *100*, 13,875–13,899.

McCormick, M. P. (1987), SAGE II: An overview, *Adv. Space Res.*, *7*(3), 219–226.

Rind, D., and X. Liao (1997), Stratospheric Aerosol and Gas Experiment II CD Rom Atlas of Global Monthly Mean Aerosols, Ozone, NO_2, Water Vapor and Relative Humidity (1985–1993), *Earth Int. J.*, *1*, 1–23.

Thomason, L. W., S. P. Burton, N. Iyer, J. M. Zawodny, and J. Anderson (2004), A revised water vapor product for the Stratospheric Aerosol and Gas Experiment (SAGE) II version 6.2 data set, *J. Geophys. Res.*, *109*, D06312, doi:10.1029/2003JD004465.

Wang, H. J., D. M. Cunnold, L. W. Thomason, J. Zawodny, and G. E. Bodeker (2002), Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res.*, *107*(D23), 4691, doi:10.1029/2002JD002418.

J. Lerner, Center for Climate Systems Research, Columbia University, 2880 Broadway, New York, NY 10025, USA.

D. Rind, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA. (drind@giss.nasa.gov)

J. Zawodny, NASA Langley Research Center, Hampton, VA 23681-0001, USA.